I.6. Improved data preprocessing

Several improvements have been made to the observation preprocessing for the Terra system. The preprocessing now handles new input data formats such as BUFR (Thorpe 1995), and creates observation datasets in the Observation Data Stream (ODS, da Silva and Redder 1995) format used by the GEOS assimilation. Additional data types such as SSM/I data are now handled. Observation filtering for high-density data types has been implemented. Improvements have also been made to the preprocessing quality control, including the implementation of the NCEP complex quality control for radiosondes. The following paragraphs describe the Terra observation preprocessing.

Conventional data (surface land, ship and buoy reports, aircraft reports, cloud motion winds, rawinsonde, dropwindsonde and pibal reports) used in the Terra system are obtained from NCEP via the Goddard Distributed Active Archive Center (DAAC). The datasets are in BUFR format. These datasets may contain quality marks assigned by the Senior Duty Meteorologist (SDM) at NCEP. The data are obtained from an early stage in the NCEP processing and thus have not been subjected to their complex quality control. NESDIS TOVS soundings are also copied from NCEP via the DAAC. These soundings will be used in the Terra assimilation until the DAO's interactive retrievals are available (Discussed in C. New Data Efforts)

In the first step of the conventional data pre-processing, the BUFR datasets are converted to NMC Office Note 29 (Keyser 1996) format. The processing then follows the steps of the GEOS-1 data assimilation system. The data are unpacked and put into common format datasets. For the first-look system, separate datasets are created for each synoptic time. The late look system concatenates the data from four synoptic times - 18 UTC of the previous day, 00, 06 and 12 UTC. All data that appear in the original data sets are kept, except those that do not have a realistic time stamp. The first look system discards reports in the wrong synoptic time, while the late look system moves these reports to the correct file.

The second step standardizes the observations for ingest into the objective analysis. Observations of quantities to be analyzed are extracted for these data sets: sea level pressure and wind, upper-air height, wind and moisture. The NESDIS temperature retrievals are converted into thicknesses. Other observations of interest, such as upper-air temperatures, are also extracted; these are carried in the assimilation as passive data types. The observations are stratified by type (e.g., surface land, surface ship) and in some instances by location.

Once the observations to be analyzed are extracted, the next step:

- ✓ keeps observations marked by the provider as 'blacklist', 'suspect' or 'purge' and marks this data with a history mark to indicate the data quality is suspect.
- ✓ eliminates observations otherwise marked as bad by the provider
- ✓ eliminates observations with grossly bad values ('range check').
- ✓ performs a hydrostatic check on rawinsonde data.
- ✓ calculates relative humidity and mixing ratio from observed pressure, temperature and dewpoint temperature.
- ✓ checks satellite profiles for completeness.

The data is written in 'repack' format, and the datasets are then converted to ODS.

NESDIS SSM/I wind and precipitable water retrievals are also copied from NCEP via the Goddard DAAC. Data is read from this BUFR file and output in ODS format. Observations, which fail a range check, are given an exclusion mark and are not used in the assimilation.

High-density observations such as the cloud motion winds from geostationary satellites must be reduced in volume before utilization in the Terra system. An observation thinning program reads each high-density data type from an ODS file and sorts the observations into latitude-longitude bins. Various algorithms can then be used to select one observation from each bin to be used in the assimilation. For the GOES cloud motion winds we are currently using bins of 2.5 degrees longitude by 2 degrees latitude and selecting the observation closest to the center of the bin to be included in the assimilation. Other observations are marked with a thinning exclusion mark before being written to the output ODS dataset.

In the next build of the Terra system, NCEP routines for preprocessing and complex quality control of rawinsonde data will be added. The NCEP PREPDATA program reformats, performs range and hydrostatic checks and applies a time window to the data. The DAO version of NCEP's PREVENTS program interpolates the DAO forecast first guess on sigma levels to observation locations, and flags observations with suspect pressure values. The NCEP CQCHT program (Collins 1991,1997,1998, Collins and Gandin 1990, 1992, 1995) performs a variety of checks on rawinsonde mass and temperature data and flags and can sometimes correct bad observations. The CQCHT checks height data and both mandatory and significant level temperature data. It calculates observation increments (observed minus background values) and performs lapse rate checks, increment checks, horizontal and vertical residual checks, hydrostatic residual checks, and baseline checks. The RADCOR radiation correction (Julian, 1991) is not implemented in the Terra system.

The 'prepbufr' file written by the CQCHT is then converted to ODS format. This ODS conversion program also does the following:

- ✓ calculates relative humidity and mixing ratio from observed pressure, temperature and dewpoint temperature.
- ✓ converts NCEP quality marks to DAO values; 'blacklist', 'suspect', 'purge', 'CQC corrected' values are regarded as suspect and assigned a history mark. Other observations flagged as 'bad' are assigned an exclusion mark.
- ✓ duplicate observations are given an exclusion mark.
- ✓ observations failing a range check or a hydrostatic check are given an exclusion mark.
- ✓ moisture observations calculated from suspect values are assigned a history mark.

Rawinsonde observations will still be processed by the original preprocessing routines as described above, but the assimilation will only read rawinsonde data from the separate raob ODS file.